Review of Marketing Science Working Papers

Volume 1, *Issue* 3 2002 *Working Paper* 4

A Flexible Class of Purchase Incidence Models

Trichy V. Krishnan* Seethu Seetharaman †

*Jesse H. Jones Graduate School of Management, Rice University
†John M. Olin School of Business

Review of Marketing Science Working Papers is produced by The Berkeley Electronic Press (bepress). http://www.bepress.com/roms

A FLEXIBLE CLASS OF PURCHASE INCIDENCE MODELS

Trichy V. Krishnan[1]

Seethu Seetharaman[2]

November 17 2000

[1] Assistant Professor of Marketing, Jesse H. Jones Graduate School of Management, Rice University, Houston, TX 77005, e-mail: krishnan@rice.edu

[2] Assistant Professor of Marketing, John M. Olin School of Business, Washington University, Campus Box 1133, One Brookings Drive, St. Louis, MO 63130-4899, e-mail: seethu@olin.wustl.edu

A FLEXIBLE CLASS OF PURCHASE INCIDENCE MODELS

Abstract

Purchase incidence models estimated on household scanner panel data typically assume the household's *decision interval* to be one week. However, it is well known in the econometrics literature that discrete-time models are highly sensitive to the assumed time interval of decision-making. In this study we investigate the consequences of endogenizing the household's decision interval, instead of restricting it to be one week. We characterize the household's random utility maximization problem, and therefore its purchase likelihood function, as a function of the household's decision interval. Such a flexible purchase incidence model is then used to explicitly estimate households' decision intervals in addition to their response to marketing activity and their baseline hazard functions. The proposed model of purchase incidence not only nests traditionally used *choice models* (such as the binary logit model) and *hazard models* (such as the discrete hazard model), but also allows for a gamut of more flexible parametric specifications. We estimate the proposed model across four category-level scanner panel datasets and find that the traditional assumption of restricting the household's decision interval to be one week may be too restrictive. We find that households are not only quite heterogeneous in their decision intervals but often have decision intervals longer than a week. From a managerial perspective, we show that estimated price elasticities are systematically understated if one does not allow for the effects of decision intervals. We demonstrate, using a fourth product category, that the results obtained from the category-level analyses generalize to the context of a full model of purchase incidence and brand choice.

Key words: Decision intervals, Purchase incidence models, Choice models, Logit, Hazard.

Introduction

Random utility models have a rich history in Marketing. These models have typically been used to characterize a household's decision of whether to buy a particular product during a shopping trip (also called the *purchase incidence* decision), and contingent on a decision to buy which of several available brands to buy (also called the *brand choice* decision). Examples of empirical studies that have estimated these two purchase decisions using a random utility framework are Gupta (1988), Chiang (1991), Chintagunta (1993) etc.

The attractiveness of employing random utility models in these contexts lies in the fact that these models stem from economic theory i.e. they are derived from a theory of rational utility maximization on the part of the household (McFadden 1986). Purchases incidence models that utilize the random utility framework have identified two main *drivers* of the purchase incidence decision i.e. variables that influence the household's utility for a product. These variables are the following:

1. <u>Marketing mix</u>: This stands for price and promotional activity associated with the product e.g. shelf price, store displays, newspaper feature advertisements etc. This information is available in conventional scanner panel data.
2. <u>Product inventory</u>: This stands for the amount of product in stock at home when the household undertakes the shopping visit. Since this information is not recorded in conventional scanner panel datasets, this effect is typically modeled using either an imputed inventory variable or some function of time since last purchase in the household's utility function.

Two types of models are useful from the point of view of characterizing the above two effects on purchase incidence: 1. *Choice models* such as the binary logit are useful to characterize the effects of the marketing mix (Bucklin and Lattin 1991), 2. *Hazard models* such as the proportional hazard are useful to characterize the effects of time since last purchase (Jain and Vilcassim 1991). *Discrete hazard* models combine the benefits of the choice and hazard approaches in a utility-consistent manner. They model the effects of marketing variables on the household's random utility for the product in the same way as a choice model, and the effects of time since last purchase using a step-function hazard (Jain and Vilcassim 1994, Wedel et al. 1995).

Purchase incidence models, including discrete hazard models, have largely assumed that the time interval of household decision-making for the product is *one week* i.e. each household is assumed to contemplate[1] purchase of the product once every week. This assumption is (implicitly) motivated by two reasons: one, households tend to visit stores in weekly intervals; two, marketing variables of products change from one week to another.

Even if a household visits the store every week, they are not likely to contemplate the purchase of a given product category on each visit. For example, a household may not actively consider a ketchup purchase during its store visit this week either on account of having adequate inventory at home or on account of other product categories being more "salient" in the context of this week's consumption needs. However, the same household may actively consider a ketchup purchase the next time they visit the store on account of having changed consumption circumstances. Suppose the household ends up *not* purchasing ketchup in both weeks, one must recognize that the second no-purchase is a consequence of the household deciding not to buy after active consideration, while the first no-purchase is simply a consequence of the household not considering the ketchup purchase at all! If a purchase incidence model does not distinguish between these two types of store visits, the estimated effects of marketing variables on the household's choices is likely to be distorted.

We will refer to the time interval between two successive store visits when a household actively considers whether or not to purchase the product category as the household's *decision interval*. Suppose households A and B visit the grocery store in weekly intervals. However, suppose A's decision interval for ketchup is two weeks, while B's decision interval is one week. If one then observes a string of two no-purchases for each household in the product category over a period of two weeks (i.e. two store visits), one arrives at different conclusions about each household. The string of two no-purchases for A is a consequence of the household considering once during the two-week interval whether or not to buy the product and deciding not to buy. However, the string of two no-purchases for B is a consequence of the household considering twice during the two-week interval whether or not to buy the product and deciding not to buy on both occasions. Not accounting for such differences across households and treating each store visit as a similar decision opportunity for each household will make one conclude, on the basis of the observed purchase strings of the households, that both households are similarly influenced by marketing variables when in fact they are not.

Now, how does one accommodate the effects of household-specific decision intervals in purchase incidence models when these decision-intervals are in fact unobserved? This is the question we address in

[1] "Contemplating purchase," means explicitly considering, at that point of time, whether or not to purchase the product.

this paper. We endogenously estimate each household's decision interval using their observed purchases, and explain the household's string of purchases in the product category on the basis of this estimated decision-interval. We demonstrate the consequences of ignoring the effects of decision interval on the estimated price elasticities.

We will refer to the time interval between two successive store visits of a household as the household's *store visit interval*[2]. Our point is that a household's *decision interval* is in general not equal to its *store visit interval*. For product categories such as ketchup, the decision interval is likely to be greater than the store visit interval (as explained earlier). For product categories such as milk, the decision interval is likely to be equal to (or even less) than the store visit interval. For example, a household is more likely to make an unscheduled trip to the store to buy milk than to buy ketchup. In other words, a household is likely to contemplate milk purchases more frequently than ketchup purchases because milk is a more indispensable component of the household's pantry. Therefore, even if a household visits the store each week, the household's decision intervals may be vastly different for different product categories within its shopping basket. A purchase incidence model that assumes a household's decision interval in a product category to be equal to its store visit interval effectively ignores such differences across product categories.

The focus of this study, therefore, is two fold: one, we explicitly model the effects of decision intervals in purchase incidence models; two, we investigate the consequences of ignoring differences in decision intervals both across households and across product categories. Our proposed solution works on the following idea: Since households visit stores at weekly intervals and walk past a majority of product aisles in the store, their purchase likelihood must be constructed on the weekly store interval. However, since households differ in their decision intervals, their purchase likelihood for each week must be adjusted to reflect these differences. For example, longer a given household's decision interval, less likely a purchase on any given week. Not adjusting for this will overestimate the effect (or lack thereof) of marketing activities on that household that week.

From an econometric standpoint, our study is in the same spirit as Ryu (1995) who persuasively argues that model inferences obtained using the discrete hazard model are in general highly sensitive to the assumed time interval of decision-making e.g. weekly, biweekly etc. We propose one way of alleviating this concern. Unless there are theoretical prescriptions to advice empirical researchers on what time interval to use in a discrete hazard model *a priori* i.e. before looking at the data, the question of decision

[2] It is usually observed that households' store visit intervals are one week (Kahn and Morrison 1989).

interval specification can be answered only using the data. This is the approach we take, and we use the notion of decision intervals to motivate the interval specification issue. We endogenously model and estimate the effects of decision intervals, in addition to the effects of marketing variables and time since last purchase, on household purchases in the product category[3].

In this study, we propose a highly parsimonious parameterization of decision intervals within a purchase incidence framework. The proposed model flexibly allows the household's decision interval for a given product category to take any non-negative real value. A notable aspect of our proposed framework is that it nests traditionally employed choice models and hazard models of purchase incidence, while in addition allowing for more flexible specifications of decision intervals (using just one additional parameter). For example, the discrete hazard models of Jain and Vilcassim (1994) and Wedel et al. (1995) correspond to a decision interval of zero (i.e. continuous-time decision making), while the binary logit model of Bucklin and Lattin (1991) or a hazard variant thereof corresponds to a decision interval of one week.

We estimate the proposed model of purchase incidence across three different categories of packaged goods – soup, detergents and toilet tissue. We find, as expected, that the decision interval is quite heterogeneous across households in each of the three categories. Given a product category, some segments of households exhibit "logit-like" behavior (i.e. decision interval of one week), some exhibit "discrete hazard-like" behavior (i.e. continuous-time decision-making), while others exhibit behavior consistent with decision intervals greater than one week. In order to generalize these effects to a brand-choice context, we also estimate a nested logit model of brand choice and purchase incidence on a fourth product category - margarine. We find that the results obtained with the purchase incidence models generalize to the nested logit model as well. We demonstrate the consequences of ignoring the effects of decision-making intervals on the estimated marketing mix elasticities. Our study highlights the need for empirical researchers to explicitly incorporate households' decision intervals while estimating random utility models of purchase incidence and/or brand choice in a product category.

The rest of the paper is organized as follows: In the next section, we present the model and the estimation procedure. In section three, we discuss the empirical results. In section four, we conclude with a summary and directions for future research.

---

[3] Testing various exogenously pre-specified decision intervals for each household and concluding, on the basis of model fit criteria, which interval is the most appropriate for the household is another way of addressing this issue. However, since one typically deals with few hundred households in scanner panel data, testing all possible permutations of decision interval lengths across households does not appear to be practically feasible.

Model and Estimation

We outline the model formulation in two steps. First, we derive the likelihood function of a purchase incidence model that allows the household's decision interval to take any value less than or equal to the household's store visit interval. Second, we extend this model to handle decision intervals that are greater than the store visit interval.

Suppose a household undertakes a shopping trip during week t, and considers whether or not to buy a given product during the trip. This can be characterized as a binary (buy versus no buy) purchase incidence decision, and modeled in a random utility framework as follows.

$$
\begin{aligned}
U_{buy,t} &= \alpha + \mathbf{X}_t * \beta + \varepsilon_t , \\
U_{o,t} &= \mathbf{Y}_t * \gamma ,
\end{aligned}
\qquad (1)
$$

where $U_{buy,\ t}$ stands for the household's utility from buying the product within its store visit interval (i.e. week t), $U_{o,t}$ stands for the household's reservation utility[4] for the product within its store visit interval, $\mathbf{X_t}$ stands for a vector of product characteristics and b stands for the associated parameter vector, $\mathbf{Y_t}$ stands for a vector of household characteristics and g stands for the associated parameter vector, and ε_t is a random error that captures the effects of variables that are unobserved by the researcher. If one assumes ε_t to follow the logistic distribution, one obtains the *logit choice model* (McFadden 1986, Bucklin and Lattin 1991). According to the logit choice model, the household's probability of buying the product within its store visit interval is given by

$$
P_{buy,t} = \frac{e^{\alpha + \mathbf{X}_t * \beta - \mathbf{Y}_t * \gamma}}{1 + e^{\alpha + \mathbf{X}_t * \beta - \mathbf{Y}_t * \gamma}} . \qquad (2)
$$

The logit choice model can permit the household's reservation utility to change with time by allowing time or an imputed inventory measure to be a variable in the vector $\mathbf{Y_t}$. Alternatively, the logit choice model can allow the reservation utility to vary over time in the form of a *step function*. This semi-parametric approach yields what is called a *logit hazard model* (Allison 1984). The logit hazard model nests the logit choice model as a special case when the step function is constant from one discrete time period to another. It is useful to note that in order to capture the effects of product inventory these approaches allow the household's reservation utility to be a function of the *time since last purchase*. Further, these approaches assume that the household's purchase incidence decision is made based on a decision interval of one-week i.e. the week of the shopping trip.

---

[4] *Reservation utility* refers to the minimal level of utility that a product must offer in order to induce a

Suppose the decision interval is less than the store visit interval i.e. the household contemplates purchase of the product even if the household does not visit the store. This assumption is valid for product categories that explicitly drive store visits e.g. milk. When one runs out of milk, one decides whether to go to the store to replenish depleted stocks of milk (even if a regular store visit is not "scheduled" for that time). This assumption may not be valid for product categories that do not drive store visits e.g. soda. When one runs out of soda, one waits until one's scheduled next store visit to replenish depleted stocks of soda (even if this means living for a few days without soda at home). Let us first address this case of the decision interval being less than or equal to the store visit interval.

Case 1: The household's decision interval is *less than* its store visit interval

Suppose the household's store visit interval is one week. Further, suppose the household's decision interval is half a week. This means that the household's purchase likelihood in week t must be written as

Pr (purchase in week t) = 1 – Pr (no purchase in week t)

= 1 – Pr (no purchase in first half-week $t_{1/2}$) * Pr (no purchase in second half-week $t_{1/2}$).

If we do not distinguish between the first half-week and the second half-week in terms of the no-purchase likelihood, this yields

Pr (purchase in week t) = 1 – Pr (no purchase in half-week $t_{1/2})^2$.

It follows then that if the household's decision interval is n^{th} of a week (where n $\leq$ 1), the household's purchase likelihood in week t can be written as

Pr (purchase in week t) = 1 – Pr (no purchase in $t_n)^{1/n}$ (3)

(Note: For n = 1 we obtain the familiar purchase incidence model with a decision-interval of one week). Now, how does one specify Pr (no purchase in t_n)? We know that this probability must be greater than Pr (no purchase in week t), since the time-interval t_n is less than the time-interval t. For example, all else being equal, the household's probability of buying within a half-week interval must be lower than the household's probability of buying within a one-week interval. Further, the probability of buying within time interval t_n must tend to zero as n tends to zero. A simple way of accommodating this effect is to operationalize the household's reservation utility as a function of the time interval characterizing its decision interval. This can be done as follows.

$$U_{buy,t_n} = \alpha + \mathbf{X}_{t_n} * \beta + \varepsilon_{t_n},$$
$$U_{o,t_n} = \mathbf{Y}_{t_n} * \gamma + f(n), \tag{4}$$

household to purchase it.

where n stands for the decision interval (in weeks), and f (n) is a decreasing function of n with f (1)=0[5]. This captures the notion that the reservation utility gets larger for smaller decision intervals. Anticipating this effect to be *concave*[6], we operationalize f (.)= -ln (.) which yields

$$U_{buy,t_n} = \alpha + \mathbf{X}_{t_n} * \beta + \varepsilon_{t_n}, \quad U_{o,t_n} = \mathbf{Y}_{t_n} * \gamma - \ln(n), \tag{5}$$

This yields the following expression for the household's probability of not purchasing within the household's decision interval n.

$$P_{nobuy,n} = \Pr(nopurchase\ in\ t_n) = \frac{1}{1+e^{\alpha+\mathbf{X}_t*\beta-\mathbf{Y}_t*\gamma+\ln(n)}} = \frac{1}{1+ne^{\alpha+\mathbf{X}_t*\beta-\mathbf{Y}_t*\gamma}}. \tag{6}$$

From this equation we can see that $P_{nobuy,\, n}$ decreases as n increases. Substituting this equation in (3), we obtain the following purchase likelihood for the household.

$$P_{buy,t} = 1 - \left(\frac{1}{1+ne^{\alpha+\mathbf{X}_t*\beta-\mathbf{Y}_t*\gamma}}\right)^{1/n}. \tag{7}$$

Equation (7) characterizes the purchase incidence model that we propose in this study. While households' store visit intervals are assumed to be exogenous (as is commonly assumed in existing work on purchase incidence models), households' decision intervals (*n*) are allowed to be flexible and are endogenously estimated. We can readily see that for n=1, the proposed model reduces to the *logit choice model* (McFadden 1986). As *n* tends to zero, we get the following limiting result.

$$Lim_{n\to 0} P_{nobuy,t} = Lim_{n\to 0}\left(\frac{1}{1+ne^{\alpha+\mathbf{X}_t*\beta-\mathbf{Y}_t*\gamma}}\right)^{1/n} = e^{-e^{\alpha+\mathbf{X}_t*\beta-\mathbf{Y}_t*\gamma}}. \tag{8}$$

The purchase and no-purchase probabilities are therefore given by the following equations.

$$P_{buy,t} = 1 - e^{-e^{\alpha+\mathbf{X}_t*\beta-\mathbf{Y}_t*\gamma}}, \quad P_{nobuy,t} = e^{-e^{\alpha+\mathbf{X}_t*\beta-\mathbf{Y}_t*\gamma}}. \tag{9}$$

This is also called the *extreme value choice model* (Heckman 1996). If Y_t contains a semi-parametric function of time, more specifically a *step function* of time, this is equivalent to the discrete *hazard models* of Jain and Vilcassim 1994 and Wedel et al. 1995, hereafter referred to as the *extreme value hazard model.* In these models, one allows α to be a step-function of time. For example, α_1 and α_2 would stand for intercepts corresponding to the first and second weeks since last purchase respectively. A plot of α_t as a function of *t* is referred to as the *baseline hazard*.

It is useful to define our nomenclature at this point. We will henceforth refer to the step function

---

[5] If n=1, we must obtain the traditional random utility model of purchase incidence that corresponds to a decision interval of one week.

[6] That is, f (n) decreases at a slower rate as n increases.

of time (that captures the effects of product inventory) as the *baseline hazard*[7]. Purchase incidence models that do not include the baseline hazard will be referred to as *choice models* e.g. logit choice model, extreme value choice model etc. Purchase incidence models that include the baseline hazard will be referred to as *hazard models* e.g. logit hazard model, extreme value hazard model etc.

Using a single parameter (n), we have proposed a purchase incidence model that not only captures the effects of a household's interval of decision-making, also called its decision interval, but also nests previously employed choice and hazard models in the literature. However, we have so far assumed that the decision-interval is less than or equal to the household's store visit interval. What if the decision interval is greater than the store visit interval? We will visit this question next.

<u>Case 2: The household's decision interval is *greater than* its store visit interval</u>

Suppose the household's store visit interval is one week (as before). However, suppose the household's decision interval is two weeks i.e. n=2. What does this mean? It means that this household does not contemplate purchase of the product every time they visit the store. One possible reason for this behavior could be that this household has a purchase cycle for this product that is longer than the household's store visit cycle so that each store visit is not necessarily a decision-making opportunity to the household. One can still write the household's purchase likelihood in week t as in equation (7), except that we now allow for the possibility that n > 1. That is,

$$P_{buy,t} = 1 - \left(\frac{1}{1 + ne^{\alpha + \mathbf{X}_t * \beta - \mathbf{Y}_t * \gamma}} \right)^{1/n}, \tag{10}$$

where $n \geq 1$. As n gets larger, this probability becomes smaller. In other words, longer the decision-interval for a household, less likely the household is to buy in any arbitrary week t (see Figure 1). For example, suppose a household contemplates the purchase of ketchup only once in two months. In that case, it is safe to assume that the household's probability of purchasing ketchup in a given week is quite small compared to, say, the household's probability of purchasing milk that week. Of course, if one knew the weeks during which the household considers whether or not to buy ketchup, the likelihood function for the household must be constructed based on those weeks only. In the absence of such information, the best one can do is to revise downward the household's purchase likelihood for ketchup in any given week.

Purchase incidence models that condition a household's product purchases on the household's store visits assume that each store visit presents an equal decision opportunity for all product categories.

[7] The *baseline hazard* refers to the household's probability of purchasing the product as a function of time, ignoring the effects of marketing variables.

Our point is that a store visit presents a greater decision opportunity for some product categories (e.g. those that are purchased more frequently) than others (e.g. those that are purchased infrequently). This is exactly what our model allows for using n > 1. To the extent that the decision-interval is reflective of the household's average purchase cycle in the product category, it may be proportional to the household's average inter-purchase time, for example. It will also pick up the effects of other phenomena (such as promotional patterns in other product categories within the household's shopping basket) that influence the interval of household decision-making. For example, a household may consider whether or not to purchase ketchup if most of the other goods in the store (called the "composite good") are on deal, which frees up some money that week for discretionary spending on ketchup. To the extent that the composite good is collectively discounted once every *n* weeks, the decision interval may reflect this time interval. Our contention is that households differ in terms of what time interval is relevant for their decision-making in the product category, and not recognizing these differences will lead to distorted inferences about their response to marketing activities[8].

We combine cases 1 and 2 using the following model.

$$P_{buy,t} = 1 - \left(\frac{1}{1 + ne^{\alpha + \mathbf{X}_t * \beta - \mathbf{Y}_t * \gamma}} \right)^{1/n}, \qquad P_{nobuy,t} = \left(\frac{1}{1 + ne^{\alpha + \mathbf{X}_t * \beta - \mathbf{Y}_t * \gamma}} \right)^{1/n}, \tag{11}$$

where $0 \leq n < \infty$. This is our proposed model. It is useful to note here that this model is different from the hazard version of the binary logit model (that has been frequently used to model purchase incidence) in *two* ways: one, it uses $ne^{\alpha + \mathbf{X}_t * \beta - \mathbf{Y}_t * \gamma}$ instead of $e^{\alpha + \mathbf{X}_t * \beta - \mathbf{Y}_t * \gamma}$ in the denominator, where *n* is the household's decision interval (in weeks); two, it has the exponent (1/n). Larger the value of n, smaller the household's probability of purchasing the product in any arbitrary week t (see Figure 1). If n = 1 this specification reduces to the binary logit model, which assumes that the decision interval is equal to the store visit interval.

Some researchers have (implicitly) recognized the role of decision-intervals by arguing that one goes from a logit choice model to a discrete hazard model if one assumes *continuous-time decision-making* (Allison 1984). This is very consistent with our proposal that the discrete hazard model corresponds to n = 0 i.e. decision-interval of length zero, which is equivalent to continuous-time decision-making. However, unlike some researchers who take the view that continuous-time models are generally

[8] In fact, we later demonstrate that this is indeed the case by comparing price elasticities across model specifications.

to be preferred since inferences from discrete-time models crucially depend on the chosen time-interval, we take the broader view that one can both determine whether a continuous-time or discrete-time model is warranted, and then estimate the "correct" time-interval based on the empirical data. We propose a parsimonious approach to endogenously estimate the time-interval of household decision-making, without exogenously assuming it to be either one week or the household-specific store visit interval.

To reiterate, our proposed model of purchase incidence parsimoniously nests previously used purchase incidence models in the literature. The following possibilities arise.

1. n = 0 corresponds to the *extreme value hazard model* (Jain and Vilcassim 1994, Wedel et al. 1995).
2. n = 0 and no baseline hazard corresponds to the *extreme value choice model* (Heckman 1996).
3. n = 1 corresponds to the *logit hazard model* (Allison 1984).
4. n = 1 and no baseline hazard corresponds to the *logit choice model* (McFadden 1986, Bucklin and Lattin 1991).
5. 0 < n < 1 corresponds to a hazard model with decision interval less than one week.
6. 0 < n < 1 and no baseline hazard corresponds to a choice model with decision interval less than one week.
7. n > 1 corresponds to a hazard model with decision interval greater than one week.
8. n > 1 and no baseline hazard corresponds to a choice model with decision interval greater than one week.

Although the above eight parametric possibilities exist to model purchase incidence behavior, marketing researchers have typically used only models 1-4 in previous work. This study investigates the entire gamut of possible parametric specifications in order to understand which best characterize purchase incidence behavior of households. The study's main contribution lies in its investigation of the role of the decision interval (*n*) in a household's purchase incidence decision, an issue that has not been addressed so far in the literature (and is addressed in models 5-8 above). The proposed model is *parsimonious* in that it allows us to directly estimate the decision interval of the household on the basis of a *single* parameter (*n*). In Figure 1, we illustrate the effects of the parameter *n* on the household's purchase likelihood for a given week t. The likelihood function at the household-level can be written as shown below.

$$L_h = \prod_{t=1}^{N_h} \left(P_{h,buy,t}\right)^{\delta_t} * \left(P_{h,nobuy,t}\right)^{1-\delta_t}, \tag{12}$$

where the subscript h is used to qualify household h, N_h stands for the number of store visits corresponding to household h, and δ_t is an indicator variable that takes the value of 1 if the product is purchased during visit t and 0 otherwise. Finally, we incorporate *unobserved heterogeneity* in the model by allowing the parameters to be distributed according to a multivariate, discrete distribution across households (Kamakura & Russell. 1989). This yields the following sample likelihood function.

$$L = \prod_{h=1}^{H} \left[\sum_{s=1}^{S} (\pi_s * L_{h,s}) \right], \tag{13}$$

where S refers to the number of mass points characterizing the multivariate discrete distribution, π_s refers to the densities corresponding to these mass points, and $L_{h,\,s}$ stands for the likelihood function of household h computed using the parameter vector corresponding to mass point s. Note that this heterogeneity specification allows the decision interval (n) to be heterogeneous across households. This allows for the possibility that while some households may be adequately characterized by a logit choice structure, others may be characterized by a logit hazard structure and so on.

This completes our formulation of the proposed model of purchase incidence. The attractiveness of this model lies in the fact that it not only nests previously proposed purchase incidence models in the literature, but also allows for arbitrary decision intervals. Although previously proposed purchase incidence models have investigated the effects of the *time since last purchase*, no attention has been paid to understanding the effects of decision intervals, which pertains to, among other things, *time until next purchase*. Addressing this is the critical contribution of this study.

Although the proposed model pertains to the purchase incidence decision, it is fairly straightforward to extend the proposed model to accommodate the household's brand choice decision as well. This is done in a *nested logit* framework as follows (see Ben Akiva and Lerman 1985 for details).

$$P_{nobuy,t} = \left(\frac{1}{1 + ne^{a + IV * g_1 - \mathbf{Y}_t * g_2}} \right)^{1/n},$$

$$P_{buy,t} = \left\{ 1 - \left(\frac{1}{1 + ne^{a + IV * g_1 - \mathbf{Y}_t * g_2}} \right)^{1/n} \right\} * \frac{\exp(\mathbf{X}_{jt} * b)}{\sum_{k=1}^{K} \exp(\mathbf{X}_{kt} * b)}, \tag{14}$$

where $\mathbf{X_{jt}}$ stands for the vector of marketing variables characterizing brand j at time t, b stands for the corresponding vector of coefficients, K stands for the number of brands in the product category, and IV stands for the *inclusive value* variable given by

$$IV = \ln \sum_{k=1}^{K} \exp(\mathbf{X}_{kt} * \beta), \tag{15}$$

and γ_1 stands for the inclusive value coefficient. We cannot offer a strict utility-based view of this nested logit specification[9]. However we estimate this model only to illustrate the benefits of modeling decision intervals in choice models that account for multiple decisions.

Empirical Results

Data

We employ A.C. Nielsen's scanner panel data on household purchases in four different categories of packaged goods: canned soup, laundry detergent, toilet tissue and stick margarine. We use the first three datasets to estimate the proposed purchase incidence model at the category level, and use the fourth dataset (i.e. margarine) to estimate a full model of purchase incidence and brand choice. These datasets cover a period of two years from January 1985 to January 1987. For each product category, we pick only those households that buy a single brand of the product on more than 80% of their purchase occasions. This is done to skirt data imputation issues. Specifically, for those weeks when a household undertakes a visit to the grocery store but does not buy the product, we use the marketing variables of the household's "favorite" brand (i.e. the brand that the household buys more than 80% of the time) as characterizing the product category. This is reasonable for households that overwhelmingly buy a single brand. On the other hand, if one picks households that switch a lot between brands, imputing marketing variables can be a challenge for those weeks when households visit stores but do not buy in the category[10]. Further, we eliminate households that are "light users" of the product (less than 4, 7 and 5 purchases over the study period for soup, tissue and detergents respectively). Descriptive statistics pertaining to the four data sets are provided in Table 1.

For soup, our household selection procedure yields a sample of 42 households making a total of 4326 shopping visits in the category, with an average inter-purchase time of 5.6 weeks. For detergents, our household selection procedure yields a sample of 73 households making a total of 7592 shopping visits in the category, with an average inter-purchase time of 8.3 weeks. For tissue, our household selection procedure yields a sample of 181 households making a total of 19367 shopping visits in the category, with an average inter-purchase time of 5.7 weeks.

For margarine, our household selection procedure yields a sample of 202 households making a total of 25250 shopping visits in the category, with an average inter-purchase time of 6.1 weeks. The

[9] We thank the area editor for alerting us to this issue.

[10] One method is to compute current averages of marketing variables across all UPCs ever bought by the household (Manchanda et al. 1999).

largest brand is Blue Bonnet, with a conditional market share of 46 %. There is more display and feature activity in this category than in the other three categories.

Benchmark model

We estimate the proposed model of purchase incidence across three *category-level* datasets i.e. soup, laundry detergents and toilet tissue. We benchmark our results against the *logit hazard model* of purchase incidence. This benchmark model is chosen for two reasons: one, this comparison allows us to explicitly examine the consequences of restricting households' decision intervals to be one week; two, we find that the logit hazard model empirically outperforms the extreme value hazard model, the logit choice model and the extreme value choice model in terms of model fit and prediction[11], and thus serves as the "best" benchmark for the proposed model. It is useful to reiterate here that the benchmark (i.e. logit hazard) model is nested within our proposed model of purchase incidence. An explicit empirical comparison of the two models is carried out to assess both the improvement in model fit and the consequences of ignoring the effects of decision intervals in purchase incidence models.

Variables

The variables included in the vector $\mathbf{X_t}$ are as follows:

1. Price ($/oz.)
2. Display (equals 1 if the product is on display, 0 otherwise)
3. Feature (equals 1 if the product is featured in a newspaper ad, 0 otherwise)

The variables included in the vector $\mathbf{Y_t}$ are as follows:

1. Shopping expenditure ($)
2. Income (thousands of $)
3. Members (i.e. family size)

We employ eight time dummies for the baseline hazard (i.e. estimate the step-function $\alpha_1,\ldots,\alpha_8$ in addition to a base intercept α_o).

Model fits

We show the goodness of fit of the proposed model of purchase incidence and the logit hazard model across the four product categories in Table 2. This table reports three measures of model fit:

[11] These results are available from the authors.

1. The value of the log-likelihood function at estimated parameter values (*LL*).
2. Schwarz Bayesian Criterion (SBC), given by $-2*LL+(\ln T)*p$, where T is the total number of observations in the dataset and p is the number of parameters in the model.
3. Akaike Information Criterion (AIC), given by $-2*LL+2*p$, where p is the number of parameters in the model.

Based on the three fit criteria, we can see that the proposed model outperforms the logit hazard model for all four product categories.

Empirical findings

We report the parameter estimates in Table 3. For soup, the estimated decision intervals for the four supports of the heterogeneity distribution are 47 weeks, 11.6 weeks, 13.1 weeks and 2.8 weeks. The baseline hazard (captured by the time intercepts α_1 to α_8) is not significant, which implies that a choice model, as opposed to a hazard model, is adequate to characterize purchase incidence in the soup category.

For detergents, the estimated decision intervals for the three supports of the heterogeneity distribution are 6.4 weeks, 10.8 weeks and 0 weeks. The baseline hazard is *flat* i.e. does not exhibit any monotonic pattern over time. Since a decision interval of 0 corresponds to the extreme value hazard model, there is some evidence in favor of such a model for this dataset. Specifically, the mass of this support point is 0.28, which can be loosely interpreted to mean that the extreme value hazard model adequately characterizes 28% of detergent buyers.

For tissue, the estimated decision intervals for the three supports of the heterogeneity distribution are 1 week, 25 weeks and 0 weeks. The baseline hazard does not exhibit a monotonic temporal pattern. Since a decision interval of 1 week corresponds to the logit hazard model of purchase incidence, there is some evidence in favor of such a model for this dataset. Specifically, the mass of the support points corresponding to n=1 and n=0 are 0.10 and 0.31 respectively. This can be loosely interpreted to mean that the extreme value hazard model of purchase incidence adequately characterizes 31% of tissue buyers, while the logit hazard model of purchase incidence adequately characterizes 10% of tissue buyers. For the remaining 59% of the buyers with a decision interval of 25 weeks, neither the traditional hazard models nor their choice model counterparts are an adequate characterization of purchase behavior.

For margarine, the estimated decision intervals for the three supports of the heterogeneity distribution are 2.6 week, 2.8 weeks and 0 weeks. The baseline hazard is almost *flat*. Since the mass of the third support point is only 0.05, there is only limited evidence in the dataset in favor of an extreme value hazard model. Since we estimate a full model of purchase incidence and brand choice using the margarine dataset, the findings obtained from the category-level analyses about decision intervals being greater than one week for a majority of households generalizes to a brand-level analysis.

For the three category-level analyses, the proposed model recovers greater variation in response parameters across households. Specifically, we are able to estimate[12] one additional support for the discrete heterogeneity distribution for the proposed model as compared to the logit hazard model e.g. four supports based on the proposed model versus three based on the logit hazard model for soup, and three versus two supports for detergents and tissue. This suggests that a limited parameterization of purchase incidence behavior leads to a limited ability to recover differences across households. Since marketers are centrally interested in characterizing and exploiting differences across households while designing tailored marketing mixes for their products, this finding has compelling relevance for managers.

For both soup and tissue, the proposed model shows superior *face validity* than the logit hazard for the estimated marketing mix effects. For example, counter-intuitive signs recovered for the price and feature coefficient for one support of the heterogeneity distribution in the logit hazard model correct themselves in the proposed model. Shopping expenditure has a positive effect on purchase incidence for soup and detergents. Income has a negative effect on purchase incidence for detergents and margarine. Family size does not have a consistent effect on purchase incidence across the four categories (negative for soup, positive for margarine, insignificant for the other two categories). We present validation results based on a holdout sample in Table 4. The proposed model fares better than the logit hazard for soup, marginally better for tissue, and marginally worse for detergents. We estimated our model for a fourth product category, yogurt, for which the proposed model vastly outperformed the logit hazard on both fit and validation criteria[13].

We have demonstrated that the proposed model of purchase incidence better characterizes household behavior compared to traditionally used purchase incidence models across four different product categories. The theoretical and empirical contribution of the proposed model is that it captures, for the first time, the effects of households' decision intervals on their purchase incidence decisions. What are the substantive implications of accommodating the effects of decision intervals i.e. how will marketing managers' decisions be affected by taking decision intervals into account?

---

[12] We keep adding support points until there is no more improvement in model fit (as in Kamakura and Russell 1989)

[13] We do not report these results in this paper since the estimated price coefficients were positive!

We compute the price elasticity of demand at the *average* observed values of the explanatory variables over the study period. We compare the price elasticity obtained using the proposed model with that obtained using the logit hazard model. The results of this comparison for the four product categories are given in Table 4. From this table we can see that the logit hazard underestimates the price elasticity of demand for all categories. This suggests that conventional logit hazard models, by not explicitly modeling the effects of households' *decision intervals*, may suffer from a systematic bias in their estimated elasticities. To the extent that managers use the estimated price elasticities to design optimal prices and price promotional schedules, logit hazard models are likely to offer them poor prescriptions for such policy-making.

Next, we investigate the effect of households' product usage rates on their decision intervals. We do this by allowing the parameter n to be a linear function of the household's usage rate, which is computed as an average measure based on the household's observed purchasing activity over the study period. This effect is consistently signed *negative*[14] i.e. higher the household's usage rate, lower its decision interval. This is consistent with one of the motivations that we provided upfront for the existence of decision intervals in households' decision-making. This indicates that usage rates may have *non-linear* effects on purchase incidence by affecting decision intervals of households. To the extent that the decision interval parameter is able to capture the effects of usage rates, it can flexibly accommodate the effects of heterogeneous inter-purchase times[15] across households.

Marketers often focus on characterizing household segments in terms of behavioral response parameters so that marketing activity can be differentially tailored to each segment. Our findings about decision intervals being heterogeneous across households is of value to such customization programs if the drivers of decision intervals can be identified and then influenced by marketing activities. Our preliminary analyses indicate that demographic variables such as income and family size do not have consistent effects[16] on the decision interval across the four categories. It will be of utmost managerial interest to explicitly characterize the drivers of decision intervals.

[14] However, it is statistically insignificant at the 0.05 level for the soup category.

[15] Inter-purchase time can be interpreted as some inverse function of the usage rate.

[16] These results are available from the authors.

Conclusions

In this study, we propose a model of purchase incidence that explicitly accommodates the effects of households' decision intervals on households' purchase incidence behavior. This model nicely generalizes existing purchase incidence models using just one additional parameter. We demonstrate, using a comprehensive estimation exercise across four categories of packaged goods, that the proposed model is statistically superior to existing choice and hazard models of purchase incidence. We illustrate the adverse consequences of ignoring the effects of decision intervals while estimating purchase incidence models, using the price elasticity measure. There are several interesting directions for future research. First, it is interesting to investigate whether a household's decision interval varies over time, and if so whether it is a function of the marketing mix. This will enable marketing managers to shorten households' decision intervals if desired. Second, it is of interest to investigate whether a given household has similar decision intervals across product categories, and if so what drives such similarities. To the extent that some households have similar decision intervals across categories, it will be worthwhile to investigate the directional bias that results in the estimated price elasticity of such households. This will allow managers to determine whether such candidates are viable candidates for targeted price promotions (using targeted coupons, for example).

FIGURE 1: PURCHASE INCIDENCE PROBABILITY ($P_{buy,\ t}$) WITHIN STORE VISIT INTERVAL (t)

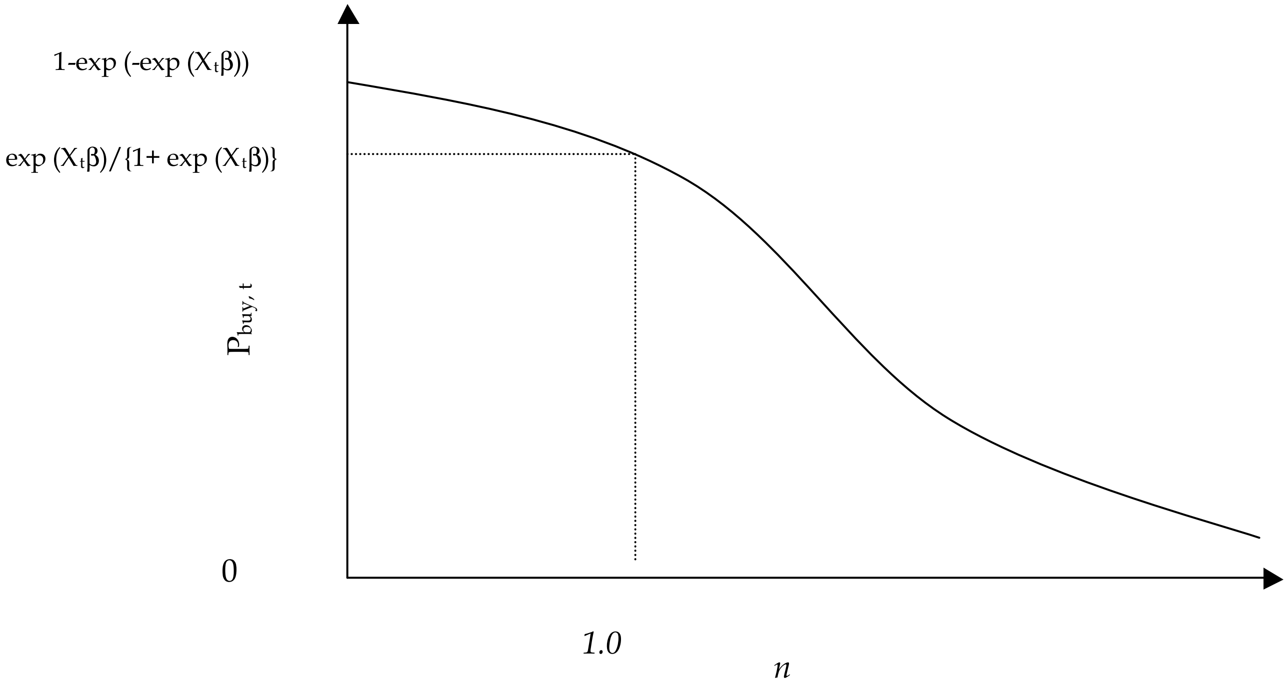

FIGURE 2: THE EFFECTS OF MARKETING VARIABLES ON THE PURCHASE INCIDENCE PROBABILITY IN A GIVEN STORE VISIT INTERVAL

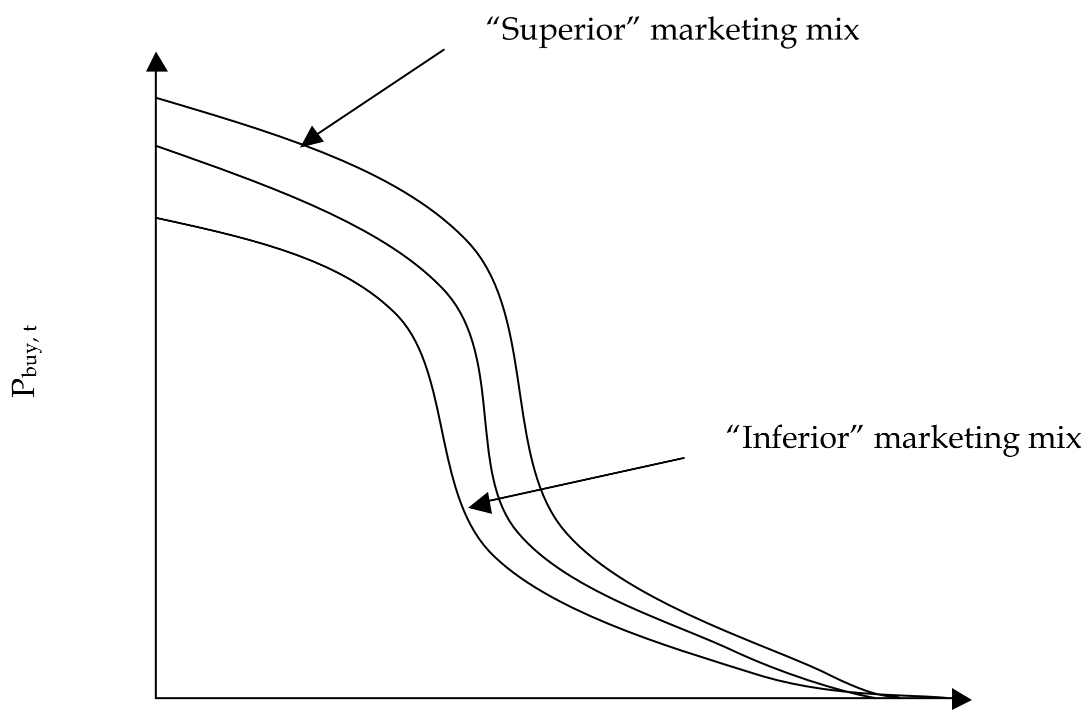

TABLE 1: Descriptive statistics

A. Category-level datasets

Product	Price ($/oz.)	Display	Feature	Avg. IPT[17]	# Observations
Soup	0.0590	0.03	0.03	5.6 weeks	4326
Detergents	0.0525	0.03	0.03	8.3 weeks	7592
Tissue	0.0920	0.13	0.21	5.7 weeks	19367

B. Brand-level dataset (Margarine)
Average inter-purchase time = 6.1 weeks, Number of observations = 25250

Brand	Price ($/oz.)	Display	Feature	Share
Blue Bonnet	0.0589	0.15	0.27	46.0 %
Parkay	0.0604	0.43	0.25	25.9 %
Imperial	0.0755	0.16	0.11	10.1 %
Fleischmann	0.1195	0.20	0.08	3.0 %
Store brand	0.0556	0.12	0.05	15.0 %

TABLE 2: Fit Results
A. Soup

Fit criterion	Proposed model	Logit Hazard
Log-likelihood	-884	-1050
SBC	2044	2300
AIC	1834	2148
# Parameters	33	24

TABLE 2 (contd.)
B. Detergents

Fit criterion	Proposed model	Logit Hazard
Log-likelihood	-1991	-2053
SBC	4223	4275
AIC	4036	4144

[17] Inter-purchase time

# Parameters	27	19

C. Tissue

Fit criterion	**Proposed model**	**Logit Hazard**
Log-likelihood	-7017	-7089
SBC	14300	14365
AIC	14088	14216
# Parameters	27	19

D. Margarine

Fit criterion	**Proposed model**	**Logit Hazard**
Log-likelihood	-9157	-9230
SBC	18719	18835
AIC	18394	18534
# Parameters	40	37

TABLE 3: Parameter Estimates (only estimates significant at the 0.05 level are reported)
A. Canned Soup

Parameter	**Proposed model (4 supports)**	**Logit Hazard (3 supports)**
α_1	Insig.	Insig.
α_2	Insig.	Insig.
α_3	Insig.	Insig.
α_4	Insig.	Insig.
α_5	Insig.	Insig.
α_6	Insig.	Insig.
α_7	Insig.	Insig.
α_8	Insig.	Insig.
α_0	0, Insig., 27.4, 23.56	0, Insig., 4.99
Price	-11.66, -3.1, -61.96, -39.03	-5.25, Insig., -8.98
Display	Insig., 3.22, 2.03, Insig.	Insig., 1.75, Insig.
Feature	Insig. for all segments	Insig. for all segments
Expend	15.18	3.44
Income	Insig.	Insig.
Members	-0.64	-1.26
N	47, 11.6, 13.1, 2.8	1

Support prob.	0.41, 0.24, 0.24, 0.11	0.18, 0.31, 0.51
LL	-884	-1050

TABLE 3 (contd.)
B. Laundry Detergents

Parameter	**Proposed model (3 supports)**	**Logit Hazard (2 supports)**
α_1	22.64	10.21
α_2	23.74	11.01
α_3	24.23	11.37
α_4	24.03	11.20
α_5	23.72	10.98
α_6	23.93	11.19
α_7	23.80	11.14
α_8	23.99	11.09
α_0	0, -25.5, -14.5	0, -12.76
Price	-7.29, -0.19, -3.29	-3.54, -0.17
Display	3.01, 2.03, 1.39	1.25, 0.89
Feature	7.40, 3.24, Insig.	1.13, 1.13
Expend	2.80	0.85
Income	-0.08	-0.05
Members	Insig.	0.07
N	6.4, 10.8, 0	1
Support prob.	0.12, 0.68, 0.28	0.37, 0.63
LL	-1991	-2053

TABLE 3 (contd.)
C. Toilet Tissue

Parameter	Proposed model (3 supports)	Logit Hazard (2 supports)
α_1	-1.28	-1.21
α_2	-1.17	-1.11
α_3	-1.42	-1.31
α_4	-1.36	-1.19
α_5	-1.71	-1.47
α_6	-1.78	-1.41
α_7	-2.04	-1.57
α_8	-3.18	-2.42
α_0	0, 1.92, Insig.	0, Insig.
Price	-0.46, Insig., -0.19	-0.56, 0.50
Display	1.81, 1.76, 0.62	0.78, 0.61
Feature	3.41, Insig., -0.36	-0.19, -0.27
Expend	Insig.	Insig.
Income	Insig.	Insig.
Members	Insig.	Insig.
N	1, 25, 0	1
Support prob.	0.10, 0.59, 0.31	0.79, 0.21
LL	-7017	-7089

TABLE 3D
Brand choice parameters

Parameter	Proposed model (3 supports)	Logit Hazard (3 supports)
$\alpha_{BlueBonnet}$	5.89, 4.16, -0.40	6.20, 4.20, -0.36
α_{Parkay}	6.08, 3.78, -1.00	6.37, 3.47, 0.11
$\alpha_{Imperial}$	6.71, 5.73, -1.70	7.20, 6.00, -0.58
$\alpha_{Fleischmann}$	19.38, 23.56, 3.41	19.39, 24.00, 5.49
$\alpha_{StoreBrand}$	0	0
Price	-3.02, -4.03, -0.55	-3.19, -4.20, -0.85
Display	-1.58, -1.65, -3.12	-1.59, -1.62, -2.06
Feature	-1.97, -2.18, -0.88	-1.96, -1.97, -1.19

TABLE 3 (contd.)
D. Stick Margarine

Purchase incidence parameters

Parameter	Proposed model (3 supports)	Logit Hazard (3 supports)
α_1	8.18	5.49
α_2	8.31	5.58
α_3	8.36	5.59
α_4	8.15	5.40
α_5	8.16	5.47
α_6	7.78	5.12
α_7	8.02	5.31
α_8	7.43	4.75
α_0	0, 4.60, -7.91	0, 3.19, -4.81
Inclusive value	0.81	0.57
Expend	Insig.	Insig.
Income	-0.01	Insig.
Members	0.12	0.10
N	2.55, 2.75, 0	1
Support prob.	0.33, 0.62, 0.05	0.34, 0.59, 0.07
LL	-9157	-9230

TABLE 4: VALIDATION RESULTS[18]

Category	Proposed model	Logit Hazard
Soup	-117	-139
Detergents	-417	-414
Tissue	-798	-799

TABLE 5: PRICE ELASTICITY OF DEMAND (Based on 3-support solutions)

Category	Proposed model	Logit Hazard
Soup	-5.70 (1.47)	-3.2 (0.77)
Detergents	-4.76 (6.55)	-2.56 (3.12)
Tissue	-0.632 (0.82)	-0.18 (0.09)

---

[18] For a fourth product category – yogurt – the proposed model and the logit hazard had validation log-likelihoods of 1379 and 1469 respectively. The results are available with the authors.

REFERENCES

Allison, P.D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*, SAGE Publications, SAGE University Papers Series: Number 07-046.

Ben-Akiva, M., Lerman, S. 1985. *Discrete Choice Analysis*, The MIT Press, Cambridge, MA.

Bucklin, R., Lattin, J. 1991. "A Two-State Model of Purchase Incidence and Brand Choice," *Marketing Science*, 10, 1, 24-39.

Chiang, J. 1991. "A Simultaneous Approach to the Whether, What and How Much to Buy Questions," *Marketing Science*, 10, 4, 297-315.

Chintagunta, P.K. 1993. "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households," *Marketing Science*, 12, 2, 184-208.

Heckman, J. 1996. *Lecture Notes on Panel Data*, Department of Economics, University of Chicago, Chicago, IL.

Gupta, S. 1988. "Impact of Sales Promotions on When, What and How Much to Buy," *Journal of Marketing Research*, 25, 4, 342-355.

Jain, D.C., Vilcassim, N.J. 1991. "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, 10, 1 1-23.

Jain, D.C., Vilcassim, N.J. 1994. "Estimating Household Purchase Rates for Consumer Nondurable Goods," *Applied Stochastic Models and Data Analysis*, 10, 15-26.

Kahn, B.E., Morrison, D.G. 1989. "A Note on Random Purchasing: Additional Insights from Dunn, Reader and Wrigley," *Applied Statistics*, 1.

Manchanda, P., Ansari, A., Gupta, S. 1999. "The Shopping Basket: A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18, 2, 95-114.

McFadden, D. 1986. "The Choice Theory Approach to Market Research," *Marketing Science*, 5, 4, 275-294.

Ryu, K. 1995. "Analysis of a Continuous-Time Proportional Hazard Model Using Discrete Duration Data," *Econometric Reviews*, 14, 3, 299-313.

Wedel, M., Kamakura, W.A., Desarbo, W.S., Hofstede, F.T. 1995. "Implications for Asymmetry, Nonproportionality and Heterogeneity in Brand Switching from Piece-wise Exponential Mixture Hazard Models," *Marketing Science*, 10, 1, 24-39.